Blackwell Optimal Strategies in Priority Mean-Payoff Games

Hugo Gimbert, Wieslaw Zielonka

Blackwell-Optimal Strategies in Priority Mean-Payoff Games

Hugo Gimbert
LaBRI, CNRS, Bordeaux, France
hugo.gimbert@labri.fr

Wiesław Zielonka
LIAFA, Université Paris 7 Denis Diderot, Paris, France
wieslaw.zielonka@liafa.jussieu.fr

We examine perfect information stochastic mean-payoff games – a class of games containing as special sub-classes the usual mean-payoff games and parity games. We show that deterministic memoryless strategies that are optimal for discounted games with state-dependent discount factors close to 1 are optimal for priority mean-payoff games establishing a strong link between these two classes.

1 Introduction

One of the recurring themes in the theory of stochastic games is the interplay between discounted games and mean-payoff games. This culminates in the seminal paper of Mertens and Neyman [12] showing that mean-payoff games have a value and this value is the limit of the values of discounted games when the discount factor tends to 1. Note however that optimal strategies in both games are very different. As shown by Shapley [13] discounted stochastic games admit memoryless optimal strategies. On the other hand mean-payoff games do not have optimal strategies, they have only ε-optimal strategies and to play optimally players need an unbounded memory.

The connections between discounted and mean-payoff games become much tighter when we consider perfect information stochastic games (games where players play in turns). As discovered by Blackwell [3], if the discount factor is close to 1 then optimal memoryless deterministic strategies in discounted games are also optimal for mean-payoff games (but not the other way round). Thus both games are related not only by their values but also through their optimal strategies. Blackwell's result extends easily to two-player perfect information stochastic games.

What happens if instead of mean-payoff games we consider parity games – a class of games more directly relevant to computer science [9]? In particular, are parity games related to discounted games?

It is well known that deterministic mean-payoff games and parity games are related, see [2]. The first insight that there is some link between parity games and discounted games is due to de Alfaro at al. [1]. It turns out that parity games are related to multi-discounted games with multiple discount factors that depend on the state. This should be compared with discounted games with a unique, state independent, discount factor which are used in the study of mean-payoff games.

Like in the classical theory of stochastic games, we examine what happens when the discount factors tend to 1, the idea is that in the limit we want to obtain parity games. Note that if we have several state dependent discount factors $\lambda_1, \dots \lambda_k$ then there are two possibilities to approach 1:

- we can study the iterated limit $\lim_{\lambda_1 \to 1} \dots \lim_{\lambda_k \to 1}$ when discount factors tend to 1 one after another (i.e. first we go to 1 with the discount factor λ_k associated with some group of states, when the limit is reached then we go to 1 with the next discount factor λ_{k-1} etc.,
- another possibility it to examine a simultaneous limit when all factors go to 1 at the same time but with different rates, this will be made precise in Section 4.

Angelo Montanari, Margherita Napoli, Mimmo Parente (Eds.)
Proceedings of GandALF 2010
EPTCS 25, 2010, pp. 7–21, doi:10.4204/EPTCS.25.5

The first approach is easier to handle than the second but it leads to weaker results, in particular we lose the links between optimal strategies in discounted games and optimal strategies in parity games.

We began our examinations of relations between discounted and parity games in [4, 5] where we limited ourselves to deterministic games. Already this preliminary work revealed that the natural framework for such a study goes far beyond parity games. In fact parity games are related to a very particular restricted class of discounted games and when we examine all multi-discounted games then at the limit we obtain a new natural class of games — priority mean-payoff games. This new class contains the usual mean-payoff games and parity games as special subclasses.

The next natural step is to try to extends the results that hold for deterministic games to perfect information stochastic games. In two papers [7, 6] we obtained some partial results in this direction. In [7] we considered a class of games that contains parity games but does not contain mean-payoff games. We showed that such games can be seen as an iterated limit of discounted games — a limit in a very strong sense, not only the value of the discounted games converges to the value of the parity game but also optimal strategies in one class are inherited by the class of games obtained in the limit. But these results are not satisfactory for two reasons, the class of games for which we were able to carry our study is too restrictive. This class involves some technical restrictions on discounted games, which are natural for parity games, but not so natural for discounted games. The second problem comes from the fact that [7] uses the iterated limit of discount factors and not the more interesting simultaneous limit.

In the second paper [6] we considered priority mean-payoff games in full generality, with no artificial restrictions, and we examined directly the limit with the discount factors tending to 1 with different rates rather than the iterated limit. However [6] deals only with one-player games and it examines only games values, the paper does not provide any relation between optimal strategies in multi-discounted games and optimal strategies in the priority mean-payoff games in the limit.

In the present paper we remove all restrictions imposed in [7, 6]. We consider the full class perfect information stochastic priority mean-payoff games and we show that such games are a limit of discounted games with discount factors tending to 1 with the rates depending on the priority. Not only at the limit the value of the discounted game equals to the value of the priority mean-payoff game but also optimal deterministic memoryless strategies in discounted games turn out to be optimal in the the corresponding priority mean-payoff game.

The interest in such a result is threefold.

First we think that establishing a very strong link between two apparently different classes of games has its own intrinsic interest.

Discounted games were thoroughly studied in the past and our result shows that algorithms for such games can, in principle, be used to solve parity games (admittedly all depends on how much the discount factor should be close to 1 in order that two types of games become close enough, and this remains open).

Another point concerns the stability of solutions (optimal strategies and games values) under small perturbations. When we examine stochastic games then the natural question is where the transition probabilities come from? If they come from an observation then the values of transition probabilities are not exact. On the other hand algorithms for stochastic games use only rational transition probabilities thus even if we know the exact probabilities we replace them by close rational values. What is the impact of such approximations on solutions, are optimal strategies stable under small perturbations? Usually we tacitly assume that this is the case but it would be better to be sure. Since Blackwell-optimal strategies studied in Section 4 are stable under small perturbations of discount factors (because they do not depend on the discount factor) this adds some credibility to the claim that Blackwell optimal strategies are stable for parity games.

And the last point. Blackwell invented Blackwell optimality because he was not satisfied with the

notion of optimal strategies for mean-payoff Markov decision processes. However the same can be said about parity games, we defer examples to the final section.

The paper is organized as follows. In Section 2 we introduce stochastic games in general, we define the notions of value and optimal strategies. Section 3 we examine discounted games. The main result in this section shows that if discount factors are close to 1 then optimal strategies stabilize (Blackwell optimality). In Section 5 we introduce the class of priority mean-payoff games — this is the principal class of games examined in this paper. Parity games and mean-payoff games are just very special subclasses of this class. In Section 6 we prove the main result of the paper stating that deterministic memoryless strategies optimal for discounted games for discount factors sufficiently close to 1 are optimal in derived priority mean-payoff games.

2 Stochastic Games with Perfect Information

Notation. In this paper $\mathbb{N}$ stands for the set of positive integers, $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, and $\mathbb{R}_+$ is the set of positive real numbers.

For each finite set X, $\mathscr{D}(X)$ is the set of probability distributions over X, i.e. it is the set of mappings $p : X \to [0,1]$ such that $\sum_{x \in X} p(x) = 1$. The support of $p \in \mathscr{D}(X)$ is the set $\{x \in X : p(x) > 0\}$.

2.1 Games and Arenas

Two players Max and Min are playing an infinite game on an arena. An arena is a tuple

$$\mathscr{A} = (\mathbf{S}, \mathbf{S}_{\mathrm{Max}}, \mathbf{S}_{\mathrm{Min}}, \mathbf{A}, (\mathbf{A}(s))_{s \in \mathbf{S}}, \delta),$$

where a finite set of states $\mathbf{S}$ is partitioned in two sets, the set $\mathbf{S}_{\mathrm{Max}}$ of states controlled by player Max and the set $\mathbf{S}_{\mathrm{Min}}$ of states controlled by player Min. For each state $s \in \mathbf{S}$ there is a non-empty finite set $\mathbf{A}(s)$ of actions available in s, $\mathbf{A} = \bigcup_{s \in \mathbf{S}} \mathbf{A}(s)$. Players Max and Min play on $\mathscr{A}$ an infinite game. If at stage $i \in \mathbb{N}_0$ the game is in a state $s_i \in \mathbf{S}$ then the player controlling s chooses an action from $\mathbf{A}(s)$ and a new state s_{i+1} is chosen with probability specified by the transition mapping δ. Transition mapping δ maps each pair (s,a), where $s \in \mathbf{S}$ and $a \in \mathbf{A}(s)$, to an element of $\mathscr{D}(\mathbf{S})$. Intuitively, if in a state s and an action a is executed then $\delta(s,a)(t)$ gives the probability that at the next stage the game is in state t. To simplify the notation we shall write $\delta(s,a,t)$ rather than $\delta(s,a)(t)$.

Throughout the paper we assume that all arenas are finite, i.e. the sets of states and actions are finite.

An arena is said to be a *one-player arena* controlled by player Max if, for every state s controlled by Min, the set $\mathbf{A}(s)$ is a singleton (in particular if all states are controlled by Max then $\mathscr{A}$ is a one-player arena controlled by Max). One-player arenas controlled by player Min are defined similarly.

A finite (resp. infinite) *play* in the arena $\mathscr{A}$ is a non-empty finite (resp. infinite) sequence of states and actions in $(\mathbf{SA})^*\mathbf{S}$ (resp. in $(\mathbf{SA})^\omega$). In the sequel "play" without any attribute will be used as a synonym of "infinite play".

2.2 Payoffs

After an infinite play player Max receives a payoff from player Min. The objectives of the players are opposite, the goal of Max is to maximize the payoff while player Min wants to minimize the payoff.

The payoff can be computed in various ways. For example in a mean-payoff game each state is labeled with a real number called the reward and after an infinite play the payoff of player Max is the

limit of mean values of the sequence of rewards. In a parity game, each state is labeled with an integer called a priority and player Max receives payoff 0 or 1 depending on the parity of the highest priority seen infinitely often. In both examples, the way the payoffs are computed is independent from the transitions rules of the game (the arena), it depends uniquely on the play.

Thus formally a payoff function is a mapping

$$u : (\mathbf{SA})^\omega \to \mathbb{R}$$

from infinite plays to real numbers.

A game is a couple $\Gamma = (\mathscr{A}, u)$ made of an arena and a payoff function. Usually we consider not a particular game but rather a class of games. In this case arenas are endowed with some additional structure, usually some labeling of states or actions (for example rewards as in mean-payoff games or priorities as in parity games) and this labeling is used to define the payoff for games in the given class.

2.3 Strategies

Playing a game the players use strategies. A *strategy* for player Max is a mapping $\sigma : (\mathbf{SA})^*\mathbf{S}_{\text{Max}} \to \mathscr{D}(\mathbf{A})$ such that for every finite play $p = s_0a_0s_1a_1\dots s_n$ with $s_n \in \mathbf{S}_{\text{Max}}$, the support of $\sigma(p)$ is a subset of the actions available in s_n, i.e. for all $a \in \mathbf{A}$, if $\sigma(p)(a) > 0$ then $a \in \mathbf{A}(s_n)$.

Strategies for player Min are defined similarly and denoted τ.

Certain types of strategies are of particular interest. A strategy is *deterministic* if it chooses actions in a deterministic way, and it is *memoryless* if it does not have any memory, i.e. choices depend only on the current state of the game, and not on the past history. Formally:

Definition 1. *A strategy σ of player $i \in \{\text{Min}, \text{Max}\}$ is said to be:*

- deterministic *if,* $\forall p \in (\mathbf{SA})^*\mathbf{S}_i$, *if* $\sigma(p)(a) > 0$ *then* $\sigma(p)(a) = 1$,
- memoryless *if,* $\forall t \in \mathbf{S}_i$ *and* $p \in (\mathbf{SA})^*$, $\sigma(pt) = \sigma(t)$.

For any finite play $p \in (\mathbf{SA})^*\mathbf{S}$ and an action $a \in \mathbf{A}$ we define the cones $\mathscr{C}(p)$ and $\mathscr{C}(pa)$ as the sets consisting of all infinite plays with prefix p and pa respectively.

In the sequel we assume that the set of infinite plays $(\mathbf{SA})^\omega$ is equipped with the σ-field $\mathscr{B}((\mathbf{SA})^\omega)$ generated by the collection of all cones $\mathscr{C}(p)$ and $\mathscr{C}(pa)$. Elements of this σ-field are called *events*. Moreover, when there is no risk of confusion, the events $\mathscr{C}(p)$ and $\mathscr{C}(pa)$ will be denoted simply p and pa.

Suppose that players Max and Min are playing accordingly to strategies σ and τ. Then after a finite play $s_0a_1\dots s_n$ the probability of choosing an actions a_{n+1} is either $\sigma(s_0a_1\dots s_n)(a_{n+1})$ or $\tau(s_0a_1\dots s_n)(a_{n+1})$ depending on whether s_n belongs to $\mathbf{S}_{\text{Max}}$ or to $\mathbf{S}_{\text{Min}}$. Fixing the initial state $s \in \mathbf{S}$ these probabilities and the transition probability δ yield the following probabilities

$$\mathbb{P}_s^{\sigma,\tau}(s_0) = \begin{cases} 1 & \text{if } s_0 = s \\ 0 & \text{if } s_0 \neq s \end{cases} \tag{1}$$

is the probability of the cone $\mathscr{C}(s_0)$,

$$\mathbb{P}_s^{\sigma,\tau}(s_0a_1\dots s_na_{n+1} \mid s_0a_1\dots s_n) = \begin{cases} \sigma(s_0a_1\dots s_n)(a_{n+1}) & \text{if } s_n \in \mathbf{S}_{\text{Max}} \\ \tau(s_0a_1\dots s_n)(a_{n+1}) & \text{if } s_n \in \mathbf{S}_{\text{Min}} \end{cases} \tag{2}$$

is the conditional probability of $\mathscr{O}(s_0a_1\dots s_na_{n+1})$ given $\mathscr{O}(s_0a_1\dots s_n)$ and

$$\mathbb{P}_s^{\sigma,\tau}(s_0a_1\dots s_na_{n+1}s_{n+1} \mid s_0a_1\dots s_na_{n+1}) = \delta(s_n,a_{n+1},s_{n+1}) \tag{3}$$

is the conditional probability of the cone $\mathscr{O}(s_0a_1\dots s_na_{n+1}s_{n+1})$ given the cone $\mathscr{O}(s_0a_1\dots s_na_{n+1})$.

Ionescu Tulcea's theorem [14] implies that there exists a unique probability measure $\mathbb{P}_s^{\sigma,\tau}$ on the measurable space $((\mathbf{SA})^\omega, \mathscr{B}(\mathbf{SA})^\omega)$ satisfying (1), (2) and (3).

2.4 Optimal strategies

Let $\mathscr{A} = (\mathbf{S}, \mathbf{S}_{\rm Max}, \mathbf{S}_{\rm Min}, \mathbf{A}, (\mathbf{A}(s))_{s\in\mathbf{S}}, \delta)$ be an arena. In the sequel we assume that all payoff mappings $u : (\mathbf{SA})^\omega \to \mathbb{R}$ are bounded and measurable (for measurability we assume that $(\mathbf{SA})^\omega$ is equipped with the σ-field described in the preceding section and $\mathbb{R}$ is equipped with the σ-field $\mathscr{B}(\mathbb{R})$ of Borel sets).

Given an initial state s and strategies σ and τ of Max and Min the expected value of the payoff u under $\mathbb{P}_s^{\sigma,\tau}$ is denoted $\mathbb{E}_s^{\sigma,\tau}[u]$.

A strategy $\sigma^\sharp$ for player Max is said to be *optimal* in a game $(\mathscr{A}, u)$ if for every state s,

$$\inf_\tau \mathbb{E}_s^{\sigma^\sharp,\tau}[u] = \sup_\sigma \inf_\tau \mathbb{E}_s^{\sigma,\tau}[u] \quad .$$

Dually a strategy $\tau^\sharp$ of player Min is *optimal* if $\sup_\sigma \mathbb{E}_s^{\sigma,\tau^\sharp}[u] = \inf_\tau \sup_\sigma \mathbb{E}_s^{\sigma,\tau}[u]$, for each state s.

In general,

$$\underline{\rm val}_s(u) := \sup_\sigma \inf_\tau \mathbb{E}_s^{\sigma,\tau}[u] \le \inf_\tau \sup_\sigma \mathbb{E}_s^{\sigma,\tau}[u] := \overline{\rm val}_s(u)$$

but when these two quantities are equal then the state s is said to have the *value* ${\rm val}_s(u) = \underline{\rm val}_s(u) = \overline{\rm val}_s(u)$, denoted also ${\rm val}_s(u,\mathscr{A})$ whenever mentioning explicitly the arena is needed. Under the hypothesis that u is measurable and bounded, Martin's theorem [11] guarantees that every state has a value. Notice however that Martin's theorem does not guarantee the existence of optimal strategies.

3 Discounted Games

Arenas for discounted games are equipped with two mappings defined on the set $\mathbf{S}$ of states. The *discount mapping*

$$\lambda : \mathbf{S} \longrightarrow [0,1)$$

associates with each state s a discount factor $\lambda(s) \in [0,1)$ and the *reward mapping*

$$r : \mathbf{S} \longrightarrow \mathbb{R} \tag{4}$$

maps each state s to a real valued reward $r(s)$.

The payoff

$$u_\lambda : (\mathbf{SA})^\omega \longrightarrow \mathbb{R}$$

for discounted games is calculated in the following way. For each play $p = s_0a_0s_1a_1s_2a_2\dots \in (\mathbf{SA})^\omega$

$$\begin{aligned} u_\lambda(p) &= (1-\lambda(s_0))r(s_0) + \lambda(s_0)(1-\lambda(s_1))r(s_1) + \lambda(s_0)\lambda(s_1)(1-\lambda(s_2))r(s_2) + \dots \\ &= \sum_{i=0}^{\infty} \lambda(s_0)\dots\lambda(s_{i-1})(1-\lambda(s_i))r(s_i) \quad . \end{aligned} \tag{5}$$

Usually when discounted games are considered it is assumed that there is only one discount factor, i.e. that there exists $\lambda \in [0,1)$ such that $\lambda(s) = \lambda$ for all $s \in \mathbf{S}$. But for us it is essential that the discount factor depends on the state.

Shapley [13] proved[1] that

Theorem 2 (Shapley)**.** *Discounted games $(\mathscr{A}, u_\lambda)$ over finite arenas admit optimal deterministic memoryless strategies for both players.*

3.1 Interpretations of discounted games

The rather obscure formula 5 can be interpreted in several ways. The usual economic interpretation is the following. The reward $r(s)$ represents the payoff that player Max receives if the state s is visited. But a given sum of money is worth more now than in the future, visiting s_i at stage i is worth $\lambda(s_1)\dots\lambda(s_{i-1})r(s_i)$ rather than $r(s_i)$ (visiting s_i is worth $r(s_i)$ only the first day). With this interpretation $\sum_{i=0}^{\infty}\lambda(s_0)\dots\lambda(s_{i-1})r(s_i)$ represents the accumulated total the payoff that player Max receives during an infinite play. However, with this interpretation it is difficult to assign a meaning to the factors $(1-\lambda(s_i))$ and such factors are essential when we consider the limit of u_λ with discount factors tending to 1.

In his seminal paper [13] Shapley gives another interpretation of (5) in terms *stopping games*. Suppose that at a stage i a state s_i is visited. Then with probability $1-\lambda(s_i)$ the nature can stop the game. Since we have assumed that $0 \le \lambda(s) < 1$ for all $s \in \mathbf{S}$, the stopping probabilities are strictly positive which implies that the game will eventually stop with probability 1 after a finite number of steps.

If the game stops in s_i then player Max receives from player Min the payment $r(s_i)$ and this ends the game. Thus here player Max receives the payoff only once, when the game stops and the payoff is determined by the last state.

If the game does not stop in s_i then there is no payment at this stage and the player controlling the state s_i chooses an action to execute.

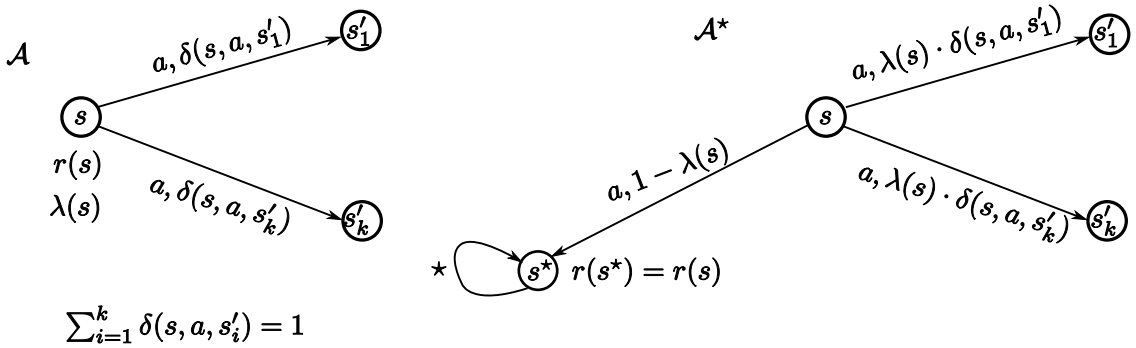
Note that $\lambda(s_0)\dots\lambda(s_{i-1})(1-\lambda(s_i))$ gives the probability that the game has not stopped in any of the states $s_0,\dots,s_{i-1}$ but it does stop in the state s_i. Since this event results in the payment $r(s_i)$, (5) represents in this interpretation the *payoff expectation* for an infinite play $s_0a_0s_1a_1s_2a_2\dots$ during the stopping game.

Another related interpretation making a direct link between discounted games and mean-payoff games is the following. We transform the discounted arena $\mathscr{A}$ into a new arena $\mathscr{A}^\star$ by attaching to each state $s \in \mathbf{S}$ a new state $s^\star$. We set $r(s^\star) = r(s)$, i.e. each new adjoined state has the same reward as the corresponding original state.

In the new arena $\mathscr{A}^\star$ we incorporate the discount factors directly into the transition probabilities. Recall that, for each state $s \in \mathbf{S}$ of the original arena $\mathscr{A}$, $\delta(s,a,s')$ was the probability of going to a state s' if an action a is executed in s. In the new arena $\mathscr{A}^\star$ this probability is set to $\delta^\star(s,a,s') = \lambda(s)\delta(s,a,s')$. On the other hand we set also $\delta^\star(s,a,s^\star) = (1-\lambda(s))$, i.e. in $\mathscr{A}^\star$ with probability $1-\lambda(s)$ the execution of a in s leads to $s^\star$ (note that for fixed a the probabilities sum up to 1).

Each new state $s^\star$ is absorbing, there is only one action available in each $s^\star$, we note it $\star$, and this action leads with probability 1 back to $s^\star$. This situation is illustrated by the following picture.

[1] In fact, Shapley considered a much larger class of stochastic games. For these games he proved that both players have memoryless optimal strategies. For perfect information games his proof yields optimal strategies that are also deterministic.



We consider the mean-payoff game played on $\mathscr{A}^\star$, i.e. the game with the payoff $u_r(s_0a_0s_1a_1\ldots) = \limsup_k \frac{1}{k+1}\sum_{i=0}^{k} r(s_i)$. Such a game played on $\mathscr{A}^\star$ ends with probability 1 in one of the starred states $s^\star$ and then the mean-payoff is simply $r(s^\star) = r(s)$. Intuitively, stopping in s with the payoff $r(s)$ in the stopping game is the same as going to $s^\star$ and looping there infinitely with the same mean-payoff $r(s^\star)$. Thus a discounted game can be seen as a mean-payoff game played on an arena where with probability 1 we end in some absorbing state. If discount factors tend to 1 then this means that, intuitively, we cut off the absorbing starred states of $\mathscr{A}^\star$.

4 Blackwell optimality

We will consider what happens if the discount factors tend to 1. The novelty in comparison with the traditional approach is that we consider the situation where discount factors of different states tend to 1 with different rates.

A *rational discount parametrization* is a family of mappings $\lambda_t = (\lambda_t(s))_{s\in\mathbf{S}}$, such that for each state s,

- $t \mapsto \lambda_t(s)$ is a rational[2] mapping of t,
- there exists $0 < \varepsilon < 1$ such that $\lambda_t(s) \in [0,1)$ for all $t \in [1-\varepsilon, 1)$ (note that since the set of states is finite we can choose the same ε for all states),
- $\lim_{t\uparrow 1} \lambda_t(s) = 1$.

A typical example of a rational parametrization is the *canonical rational discount parametrization* defined in the following way. For each state s we fix a natural number $\pi(s) \in \mathbb{N}$ called the priority of s and a positive real number $w(s) \in (0,\infty)$ called the weight of s. Then the canonical parametrization is defined as

$$\lambda_t(s) = 1 - w(s)(1-t)^{\pi(s)}, \quad \text{for } s \in \mathbf{S}, t \in \mathbb{R}. \tag{6}$$

We will consider discounted games where discount factors are given by a rational discount parametrization.

Theorem 3 (Blackwell optimality)**.** *Let us fix an arena $\mathscr{A}$ of a discounted game and let λ_t be a rational discount parametrization for $\mathscr{A}$. Let $\mathrm{val}_s(u_{\lambda_t})$ be the value of a state $s \in \mathbf{S}$ for λ_t in the game $(\mathscr{A}, u_{\lambda_t})$.*

Then there exists $0 < \varepsilon < 1$ such that, for each state s,

(1) for $t \in (1-\varepsilon, 1)$, $t \mapsto \mathrm{val}_s(u_{\lambda_t})$ is a rational function of t and

(2) if $\sigma^\sharp$ and $\tau^\sharp$ are optimal deterministic memoryless strategies for some $t \in (1-\varepsilon,1)$ then $\sigma^\sharp$ and $\tau^\sharp$ are optimal for all $t \in (1-\varepsilon,1)$.

[2]Rational in the sense that $\lambda_t(s)$ is a quotient of two polynomials of t.

In the sequel we call strategies $\sigma^\sharp$ and $\tau^\sharp$ Blackwell optimal for a rational discount parametrization λ_t if $\sigma^\sharp$ and $\tau^\sharp$ are deterministic memoryless strategies satisfying part (2) of Theorem 3.

Let us note that Theorem 3 exhibits a curious property of discounted games discovered by Blackwell [3][3]. By Theorem 2 we know that for each fixed t the discounted game with payoff u_{λ_t} has optimal memoryless deterministic strategies, but obviously such strategies depend on t. Theorem 3 asserts that for $t \in (1-\varepsilon, 1)$ the situation stabilizes and optimal deterministic memoryless strategies do not depend on t. Since Blackwell optimality is usually proved only for Markov decision processes with a unique discount factor for all states, see [10] for example, we decided to include the complete proof of Theorem 3. Note however that our proof follows closely the one used for Markov decision processes.

The proof of Theorem 3 is based on the following lemma that will be useful also in the next section.

Lemma 4. *Let $t \mapsto \lambda_t$ be a rational discount parametrization and let σ, τ be deterministic memoryless strategies. Then, for each state s, and for t sufficiently close to* 1, $\mathbb{E}_s^{\sigma,\tau}\left[u_{\lambda_t}\right]$ *is a rational function of t.*

Proof. The proof is standard but we give it for the sake of completeness. The set $\mathbb{R}^{\mathbf{S}\times\mathbf{S}}$ of functions from $\mathbf{S}\times\mathbf{S}$ into real numbers can be seen as the set of square real valued matrices with rows and columns indexed by $\mathbf{S}$. In particular $\mathbb{R}^{\mathbf{S}\times\mathbf{S}}$ is a vector space with natural matrix addition and scalar multiplication. However, matrix multiplication defines also a product on $\mathbb{R}^{\mathbf{S}\times\mathbf{S}}$, for $M, N \in \mathbb{R}^{\mathbf{S}\times\mathbf{S}}$, MN is an element U of $\mathbb{R}^{\mathbf{S}\times\mathbf{S}}$ with entries $U[s',s''] = \sum_{s\in\mathbf{S}} M[s',s]N[s,s'']$. We endow $\mathbb{R}^{\mathbf{S}\times\mathbf{S}}$ with a norm, for $M \in \mathbb{R}^{\mathbf{S}\times\mathbf{S}}$, $||M|| = \max_{s'\in\mathbf{S}} \sum_{s''\in\mathbf{S}} |M[s',s'']|$. It can be easily shown that $||MN|| \leq ||M|| \cdot ||N||$ for $M, N \in \mathbb{R}^{\mathbf{S}\times\mathbf{S}}$ and $\mathbb{R}^{\mathbf{S}\times\mathbf{S}}$ is a complete metric space for the metric induced by the norm $||\cdot||$, see Section 3.2.1 of [15] for a proof.

On the other hand, we consider also the vector space $\mathbb{R}^{\mathbf{S}}$ of functions from $\mathbf{S}$ into $\mathbb{R}$, they can be seen as column vectors indexed by states. Of course if $M \in \mathbb{R}^{\mathbf{S}\times\mathbf{S}}$ and $v \in \mathbb{R}^{\mathbf{S}}$ then $Mv \in \mathbb{R}^{\mathbf{S}}$, where $(Mv)[s] = \sum_{s'\in\mathbf{S}} M[s,s']v[s']$ for $s \in \mathbf{S}$.

We equip $\mathbb{R}^{\mathbf{S}}$ with a norm, for $v \in \mathbb{R}^{\mathbf{S}}$, $||v||_\infty = \max_{s\in\mathbf{S}} |v[s]|$. The norms on $\mathbb{R}^{\mathbf{S}\times\mathbf{S}}$ and $\mathbb{R}^{\mathbf{S}}$ are compatible in the sense that we have $||Mv||_\infty \leq ||M|| \cdot ||v||_\infty$.

Let σ, τ be deterministic memoryless strategies for players Max and Min and let λ_t be a rational discount parametrization. We define

$$\delta(s',s'') = \begin{cases} \delta(s',\sigma(s'),s'') & \text{if } s' \in \mathbf{S}_{\text{Max}}, \\ \delta(s',\tau(s'),s'') & \text{if } s' \in \mathbf{S}_{\text{Max}}, \end{cases} \quad \text{for } s',s'' \in \mathbf{S}.$$

Thus δ defines transition probabilities of the Markov chain obtained when we fix the strategies σ and τ.

In the sequel M will denote the element of $\mathbb{R}^{\mathbf{S}\times\mathbf{S}}$ defined in the following way

$$M[s',s''] = \lambda_t(s')\delta(s',s''), \quad \text{for } s',s'', \in \mathbf{S} \tag{7}$$

Let $I \in \mathbb{R}^{\mathbf{S}\times\mathbf{S}}$ be the identity matrix, i.e. $I[s',s'']$ is 1 if $s' = s''$ and 0 otherwise.

We shall show that for t close to 1 the matrix $(I-M)$ is invertible and

$$(I-M)^{-1} = \sum_{i=0}^{\infty} M^i. \tag{8}$$

First we show that the series on the right-hand side of (8) converges.

[3] In fact Blackwell [3] considered only one-player games with the same discount factor for all states.

Let $\lambda_M = \max_{s\in\mathbf{S}} \lambda_t(s)$. Then for t sufficiently close to 1 we have $||M|| \leq \lambda_M < 1$ and, for $k < l$,

$$||\sum_{i=k}^{l} M^i|| \leq \sum_{i=k}^{l} ||M||^i \leq \sum_{i=k}^{l} (\lambda_M)^i = \frac{\lambda_M^k - \lambda_M^{l+1}}{1-\lambda_M} \xrightarrow[k,l\to\infty]{} 0$$

since, by the definition of a rational discount parametrization, $0 \leq \lambda_M < 1$ for t sufficiently close to 1. Thus the series $\sum_{i=0}^{\infty} M^i$ satisfies the Cauchy condition and the convergence follows from the completeness of the norm $||\cdot||$. Now it suffices to note that

$$(I-M)^{-1} \cdot \sum_{i=0}^{k} M^i - I = M^{k+1}$$

and $||M^{k+1}|| \leq ||M||^{k+1} \leq \lambda_M^{k+1} \xrightarrow[k\to\infty]{} 0$ which yields (8).

Let $(S_i)_{i=0}^{\infty}$ be the stochastic process giving the state at stage i. Then

$$\mathbb{E}_s^{\sigma,\tau}\left[u_{\lambda_t}\right] = \mathbb{E}_s^{\sigma,\tau}\left[\sum_{i=0}^{\infty} \lambda_t(S_0)\cdots\lambda_t(S_{i-1})(1-\lambda_t(S_i))r(S_i)\right]$$
$$= \lim_{k\to\infty} \mathbb{E}_s^{\sigma,\tau}\left[\sum_{i=0}^{k} \lambda_t(S_0)\cdots\lambda_t(S_{i-1})(1-\lambda_t(S_i))r(S_i)\right] \quad (9)$$

where the second equality follows from the Lebesgue dominated convergence theorem.

Let v be an element of $\mathbb{R}^{\mathbf{S}}$ defined as

$$v[s] = (1-\lambda_t(s))r(s), \quad \text{for } s \in \mathbf{S}.$$

An elementary induction on i shows that, for $s, s' \in \mathbf{S}$,

$$\mathbb{E}_s^{\sigma,\tau}\left[\lambda_t(S_0)\cdots\lambda_t(S_{i-1})|S_0 = s, S_i = s'\right] = M^i[s,s'],$$

i.e. the entry $[s,s']$ of the i-th power of M is the expectation of $\lambda_t(S_0)\cdots\lambda_t(S_{i-1})$ under the condition that $S_0 = s$ and $S_i = s'$. This yields

$$(M^i v)[s] = \sum_{s'\in\mathbf{S}} M^i[s,s'] \cdot v[s'] = \sum_{s'\in\mathbf{S}} \mathbb{E}_s^{\sigma,\tau}\left[\lambda_t(S_0)\cdots\lambda_t(S_{i-1})|S_0 = s, S_i = s'\right] \cdot (1-\lambda_t(s'))r(s') =$$
$$\mathbb{E}_s^{\sigma,\tau}\left[\lambda_t(S_0)\cdots\lambda_t(S_{i-1})(1-\lambda_t(S_i))r(S_i)|S_0 = s\right]. \quad (10)$$

Taking the sum from $i = 0$ to k on both sides of (10) and next the limit with k tending to infinity, using (9) and (8), we obtain

$$\mathbb{E}_s^{\sigma,\tau}\left[u_{\lambda_t}\right] = ((I-M)^{-1}v)[s].$$

But the elements of the matrix $I-M$ are rational functions of t, thus Cramer's rule for matrix inversion show that $(I-M)^{-1}$ has also rational elements, and since the elements of v are also rational functions we can see that $\mathbb{E}_s^{\sigma,\tau}\left[u_{\lambda_t}\right]$ is a rational function of t.

□

Proof of Theorem 3. According to Lemma 4, and since discounted games admit optimal deterministic memoryless strategies, (1) is a consequence of (2).

We prove (2) as follows.

Let X be the set of all tuples $(q,\sigma,\tau,\sigma',\tau')$, where q is a state, σ,σ' are deterministic memoryless strategies for player Max and τ,τ' are deterministic memoryless strategies for player Min. Note that for finite arenas X is finite. Let λ_t be a rational discount parametrization and let $0<\varepsilon<1$ be such that $\lambda_t(s)\in(0,1)$ for all states s and all $t\in(1-\varepsilon,1)$.

For each $(q,\sigma,\tau,\sigma',\tau')\in X$ we consider the function $\Phi_{q,\sigma,\tau,\sigma',\tau'}:(1-\varepsilon,1)\to\mathbb{R}$ defined by:

$$t\mapsto\Phi_{q,\sigma,\tau,\sigma',\tau'}(t)=\mathbb{E}_q^{\sigma,\tau}\left[u_{\lambda(t)}\right]-\mathbb{E}_q^{\sigma',\tau'}\left[u_{\lambda(t)}\right]\ .$$

According to Lemma 4, $\Phi_{q,\sigma,\tau,\sigma',\tau'}(t)$ is a rational function of t for t sufficiently close to 1. Since a rational function can change the sign (cross the x-axis) only finitely many times there exists $\varepsilon_1=\varepsilon_1(q,\sigma,\tau,\sigma',\tau')>0$ such that the sign of $\Phi_{q,\sigma,\tau,\sigma',\tau'}(t)$ does not change in the interval $(1-\varepsilon_1,1)$. Let $\varepsilon_2=\min\{\varepsilon\}\cup\{\varepsilon_1(q,\sigma,\tau,\sigma',\tau'):(q,\sigma,\tau,\sigma',\tau')\in X\}$.

Since X is finite the minimum on the right is taken over a finite set of positive numbers and we conclude that $\varepsilon_2>0$

Let us take any $t\in(1-\varepsilon_2,1)$. Let $\sigma^\sharp$, $\tau^\sharp$ be optimal deterministic memoryless strategies in the discounted game $(\mathscr{A},u_{\lambda_t})$ (Theorem 2). Then, in particular, we have

$$\mathbb{E}_q^{\sigma,\tau^\sharp}\left[u_{\lambda_t}\right]\le\mathbb{E}_q^{\sigma^\sharp,\tau^\sharp}\left[u_{\lambda_t}\right]\le\mathbb{E}_q^{\sigma^\sharp,\tau}\left[u_{\lambda_t}\right]\tag{11}$$

for all deterministic memoryless strategies σ,τ. We can rewrite (11) as $\Phi_{q,\sigma^\sharp,\tau^\sharp,\sigma,\tau^\sharp}(t)\ge0$ and $\Phi_{q,\sigma^\sharp,\tau,\sigma^\sharp,\tau^\sharp}(t)\ge0$. However if these inequalities hold for some $t\in(1-\varepsilon_2,1)$ then we have seen that they hold for all $t\in(1-\varepsilon_2,1)$. Therefore (11) holds for all $t\in(1-\varepsilon_2,1)$. Finally Theorem 2 implies that if (11) holds for all deterministic memoryless strategies σ and τ (with fixed deterministic memoryless $\sigma^\sharp$ and $\tau^\sharp$) then it holds for all strategies σ,τ[4]. □

5 Priority mean-payoff games

In mean-payoff games the players try to optimize (maximize/minimize) the mean value of the payoff received at each stage. In such games the *reward mapping*

$$r:\mathbf{S}\longrightarrow\mathbb{R}\tag{12}$$

gives, for each state s, the payoff received by player Max when s is visited. The payoff of an infinite play is defined as the limit of the means of daily payments:

$$u_r(s_0s_1s_2\ldots)=\limsup_k\frac{1}{k+1}\sum_{i=0}^{k}r(s_i)\ ,\tag{13}$$

where we take lim sup rather than the simple limit since the latter may not exist.

We slightly generalize mean-payoff games by equipping arenas with a new mapping

$$w:\mathbf{S}\longrightarrow\mathbb{R}_+\tag{14}$$

[4]In other words, for discounted games being optimal in the class of memoryless deterministic strategies implies being optimal in the class of all strategies.

associating with each state s a *strictly positive* real number $w(s)$, the *weight* of s. We can interpret $w(s)$ as the amount of time spent in state s upon each visit to s. In this setting $r(s)$ should be seen as the payoff by a time unit when s is visited, thus the weighted mean payoff received by player Max is

$$u_{r,w}(s_0 s_1 s_2 \ldots) = \limsup_k \frac{\sum_{i=0}^{k} w(s_i) r(s_i)}{\sum_{i=0}^{k} w(s_i)} \quad . \tag{15}$$

Note that in the special case when the weights are all equal to 1, the weighted mean value (15) reduces to (13).

As a final ingredient we add to the arena a *priority mapping*

$$\pi : \mathbf{S} \longrightarrow \mathbb{N} \tag{16}$$

assigning to each state s a positive integer *priority* $\pi(s)$.

We define the *priority* of a play $p = s_0 a_0 s_1 a_1 s_2 a_2 \ldots$ as the *smallest* priority appearing infinitely often in the sequence $\pi(s_0)\pi(s_1)\pi(s_2)\ldots$ of priorities visited in p:

$$\pi(p) = \liminf_i \pi(s_i) \quad . \tag{17}$$

For any priority α, let $\mathbf{1}_\alpha : \mathbf{S} \longrightarrow \{0,1\}$ be the indicator function of the set $\{s \in \mathbf{S} \mid \pi(s) = \alpha\}$, i.e.

$$\mathbf{1}_\alpha(s) = \begin{cases} 1 & \text{if } \pi(s) = \alpha \\ 0 & \text{otherwise.} \end{cases} \tag{18}$$

Then the priority mean-payoff of a play $p = s_0 a_0 s_1 a_1 s_2 a_2 \ldots$ is defined as

$$u_{r,w,\pi}(p) = \limsup_k \frac{\sum_{i=0}^{k} \mathbf{1}_{\pi(p)}(s_i) \cdot w(s_i) \cdot r(s_i)}{\sum_{i=0}^{k} \mathbf{1}_{\pi(p)}(s_i) \cdot w(s_i)} \quad . \tag{19}$$

In other words, to calculate priority mean payoff $u_{r,w,\pi}(p)$ we take weighted mean payoff but with the weights of all states having priorities different from $\pi(p)$ shrunk to 0. (Let us note that the denominator $\sum_{i=0}^{k} \mathbf{1}_{\pi(p)}(s_i) \cdot w(s_i)$ is different from 0 for k large enough, in fact it tends to infinity since $\mathbf{1}_{\pi(p)}(s_i) = 1$ for infinitely many i. For small k the numerator and the denominator can be equal to 0 and then, to avoid all misunderstanding, it is convenient to assume that the indefinite value $0/0$ is equal to $-\infty$.)

In the sequel the couple (w, π) consisting of a weight mapping and a priority mapping will be called a *weighted priority system*.

Let us note that priority mean-payoff games are a vast generalization of parity games. In fact parity games correspond to a very particular case of priority mean-payoff games, we recover the usual parity games when we set for each state s, $w(s) = 1$ and $r(s) = 1$ if $\pi(s)$ is even and $r(s) = 0$ if $\pi(s)$ is odd.

Theorem 5. *Priority mean-payoff games over finite arenas admit optimal deterministic memoryless strategies for both players.*

Proof. The proof of Theorem 5 relies on the transfer theorem proved in [8]. This theorem states the following: if a payoff function u admits optimal deterministic memoryless strategies in all one-player perfect information stochastic games over finite arenas equipped with payoff u or $-u$, then all two-player perfect information stochastic games over finite arenas with payoff u have also optimal deterministic memoryless strategies for both players.

In [6], we proved that one-player games equipped with the payoff function $u_{r,w,\pi}$ have optimal deterministic memoryless strategy. It remains to prove the same for one-player games equipped with the payoff function $-u_{r,w,\pi}$:

$$-u_{r,w,\pi}(s_0s_1s_2\cdots) = \liminf_k \frac{-\sum_{i=0}^{k} \mathbf{1}_{\pi(p)}(s)\cdot w(s_i)\cdot r(s_i)}{\sum_{i=0}^{k} \mathbf{1}_{\pi(p)}(s_i)\cdot w(s_i)} \quad . \tag{20}$$

Let us denote $-r$ the reward mapping defined by $(-r)(s) = -r(s)$. Then,

$$u_{-r,w,\pi}(s_0s_1s_2\cdots) = \limsup_k \frac{-\sum_{i=0}^{k} \mathbf{1}_{\pi(p)}(s)\cdot w(s_i)\cdot r(s_i)}{\sum_{i=0}^{k} \mathbf{1}_{\pi(p)}(s_i)\cdot w(s_i)} \quad . \tag{21}$$

The expected values of $-u_{r,w,\pi}$ and $u_{-r,w,\pi}$ coincide on Markov chains, because in a Markov chain, the limsup in (21) is almost-surely a limit, see the proof of Theorem 7, page 8 of [6]. Since for every play, $-u_{r,w,\pi}(p) \leq u_{-r,w,\pi}(p)$, this implies that in a one-player arena, every deterministic memoryless strategy optimal for the payoff function $u_{-r,w,\pi}$ is optimal for the payoff function $-u_{r,w,\pi}$ as well, and these two games have the same values and the same deterministic memoryless optimal strategies. This completes the proof. □

6 From rationally parametrized discounted games to priority mean-payoff games

6.1 Priority mean-payoff derived from rational discount parametrization

The aim of this short subsection is to show how a rational discount parametrization induces in a canonical way a weighted priority system.

Let λ_t be a rational discount parametrization. The fact that $\lim_{t\uparrow 1}(1-\lambda_t(s)) = 0$ implies that for each state s, the function $t \mapsto 1-\lambda_t(s)$ factorizes as $g_s(t)(1-t)^{\pi(s)}$ where $\pi(s) \in \mathbb{N}$ is a positive integer constant and $t \mapsto g_s(t)$ is a rational function such that $g_s(1) \neq 0$. Moreover since $1-\lambda_t(s)$ is positive for $t \in (1-\varepsilon, 1)$, $g_s(t)$ is also positive in the same interval and by continuity of $g_s(t)$, $g_s(1) > 0$.

Now, for each state s, take $\pi(s)$ defined above as the priority of s and $w(s) := g_s(1)$ as the weight of s. We say that (w,π) defined in this way is the *weighted priority system* derived from the rational discount parametrization λ_t.

6.2 Limit of a discounted game

The following theorem establishes a remarkable link between discounted games and weighted priority mean-payoff games. Roughly speaking it shows that the latter are the limit of discounted games, the limit not only in the sense of game values (part (a)) but also the optimality of strategies is preserved in the limit.

Theorem 6. *Let $\mathscr{A}$ be a fixed arena and let $t \mapsto \lambda_t$ be a rational discount parametrization for $\mathscr{A}$. Let (w,π) be the weighted priority system derived from λ_t. Finally let $\sigma^\sharp$ and $\tau^\sharp$ be deterministic memoryless Blackwell optimal strategies for the discounted game $(\mathscr{A}, u_{\lambda_t})$.*

Then

(a) for each state s, $\lim_{t\uparrow 1} \mathrm{val}_s(u_{\lambda_t}) = \mathrm{val}_s(u_{r,w,\pi})$, where $\mathrm{val}_s(u_{\lambda_t})$ is the value of the game $(\mathscr{A}, u_{\lambda_t})$ and $\mathrm{val}_s(u_{r,w,\pi})$ is the value of the game $(\mathscr{A}, u_{r,w,\pi})$, and

(b) if $\sigma^\sharp$ and $\tau^\sharp$ are Blackwell optimal memoryless deterministic strategies for the discounted game $(\mathscr{A}, u_{\lambda_t})$ then $\sigma^\sharp$ and $\tau^\sharp$ are optimal for the priority mean-payoff game $(\mathscr{A}, u_{r,w,\pi})$

Let us note that part (a) of Theorem 6 was proved in [6] but only for one-player games[5] (Markov decision processes).

However, in [6] we were unable to establish any result linking optimal strategies for discounted games with optimal strategies of weighted priority games. Thus the main achievement of the present paper is part (b) of Theorem 6.

The following result was proved in [6] (Theorem 7 in [6]):

Lemma 7. *Let λ_t be a rational discount parametrization and let (w,π) be the derived weighted priority system. Then for each state s and for all deterministic memoryless strategies σ, τ:*

$$\lim_{t\uparrow 1} \mathbb{E}_s^{\sigma,\tau}\left[u_{\lambda(t)}\right] = \mathbb{E}_s^{\sigma,\tau}\left[u_{r,w,\pi}\right].$$

Proof of Theorem 6. We begin with part (b). Let $\sigma^\sharp$, $\tau^\sharp$ be Blackwell optimal deterministic memoryless strategies for λ_t. Let σ and τ be any deterministic memoryless strategies of players Max and Min. Then

$$\mathbb{E}_s^{\sigma,\tau^\sharp}\left[u_{\lambda_t}\right] \le \mathbb{E}_s^{\sigma^\sharp,\tau^\sharp}\left[u_{\lambda_t}\right] \le \mathbb{E}_s^{\sigma^\sharp,\tau}\left[u_{\lambda_t}\right].$$

Taking the limit with $t \uparrow 1$ we get by Lemma 7

$$\mathbb{E}_s^{\sigma,\tau^\sharp}\left[u_{r,w,\pi}\right] \le \mathbb{E}_s^{\sigma^\sharp,\tau^\sharp}\left[u_{r,w,\pi}\right] \le \mathbb{E}_s^{\sigma^\sharp,\tau}\left[u_{r,w,\pi}\right],$$

which shows that $\sigma^\sharp$ and $\tau^\sharp$ are optimal in the class of deterministic memoryless strategies. But Theorem 5 implies that for priority mean-payoff games strategies optimal in the class of deterministic memoryless strategies are optimal also when all strategies are allowed. This terminates the proof of (b).

Obviously (a) follows from (b) and from Lemma 7.

□

7 Optimal but not Blackwell optimal strategies

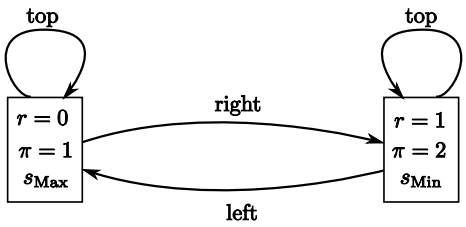


Figure 1: A parity game. Player Max has two deterministic memoryless optimal strategies but only one of them is Blackwell optimal.

Theorem 6 stated that Blackwell optimal strategies are also optimal for priority mean-payoff games. The converse is not true, the notion of Blackwell optimal strategies is strictly more restrictive.

[5] In fact, [6] shows that the convergence of game values holds not only for rational parametrizations but for any "reasonable" parametrization of discount factors.

We illustrate this with the game presented in Figure 1. Here we have two states $s_{\mathrm{Max}}, s_{\mathrm{Min}}$ controlled respectively by players Max and Min. Both states have the same weight 1 which is omitted. The left state has priority $\pi = 2$ and reward $r = 0$, the right state has priority $\pi = 1$ and reward $r = 1$, thus essentially this is the usual parity game with two priorities. Both players have two deterministic memoryless strategies. The optimal strategy for player Min is to take action "left". With this strategy state s_{Max} with priority 1 is visited infinitely often and since this is the minimal priority in this games the resulting payoff will 0 whatever the strategy of player Max. Player Max can play "top" or "right", in both cases if player Min uses the strategy described above the payoff is 0 thus both strategies are optimal for Max.

Now let us consider the associated discounted game with the canonical parametrization. Thus the discount factor of s_{Max} is $\lambda_t(s_{\mathrm{Max}}) = 1-(1-t)^{\pi(s_{\mathrm{Max}})} = t$ while the discount factor for s_{Min} is $\lambda_t(s_{\mathrm{Min}}) = 1-(1-t)^{\pi(s_{\mathrm{Min}})} = 1-(1-t)^2$. For player Min the optimal strategy is still to always play "left". For player Max the strategies "right" and "top" are now different. For example if we start from s_{Max} then playing "top" will result in payoff 0 since we will visit only the state s_{Max} with reward 0. On the other hand playing "right" we will visit infinitely often the state s_{Min} with a positive reward, thus for discounted games playing "right" is strictly better for Max than playing "top" and the strategy where Max plays "right" is the only Blackwell optimal strategy.

The main motivation behind Blackwell optimal strategies comes from the following observation (due to Blackwell). Consider a mean-payoff game controlled completely by player Max and suppose that there are only two possible infinite plays. The first play begins with a long but finite sequence of rewards 0 followed by an infinite sequence of rewards 1. The mean payoff for such history is 1, the initial sequence of 0 does not count on the limit. Consider now the second play which is an infinite sequence of rewards 1, without any 0. Here also the mean payoff is also 1. Thus player Max is indifferent between two histories. But from the point of view of Maximizer clearly the second history is better than the first one, one prefers to have the reward 1 each day rather than to begin with the reward 0. This difference is captured by Blackwell optimality.

References

[1] L. de Alfaro, T. A. Henzinger & R. Majumdar (2003): *Discounting the Future in Systems Theory*. In: *ICALP 2003, LNCS* 2719, Springer, pp. 1022–1037.

[2] H. Björklund, S. Sandberg & S. Vorobyov (2004): *Memoryless determinacy of parity and mean payoff games: a simple proof. Theor. Computer Science* 310, pp. 365–378.

[3] D. Blackwell (1962): *Discrete dynamic programming. Annals of Mathematical Statistics* 33, pp. 719–726.

[4] H. Gimbert & W. Zielonka (2006): *Deterministic priority mean-payoff games as limits of discounted games*. In: *ICALP 2006, LNCS* 4052, part II, Springer, pp. 312–323.

[5] H. Gimbert & W. Zielonka (2007): *Applying Blackwell optimality: priority mean-payoff games as limits of multi-discounted games*. In: *Logic and Automata. Hostory and Perspectives., Texts in Logic and Games* 2, Amsterdam University Press, pp. 331–355.

[6] H. Gimbert & W. Zielonka (2007): *Limits of multi-discounted Markov decision processes*. In: *LICS 2007*, IEEE Computer Society Press, pp. 89–98.

[7] H. Gimbert & W. Zielonka (2007): *Perfect information stochastic priority games*. In: *ICALP 2007, LNCS* 4596, Springer, pp. 850–861.

[8] Hugo Gimbert & Wieslaw Zielonka (2010): *Pure and Stationary Optimal Strategies in Perfect-Information Stochastic Games*. Technical Report HAL 00438359, HAL archives ouvertes. Available at `http://hal.archives-ouvertes.fr/hal-00438359/en/`.

[9] E. Grädel, W. Thomas & T. Wilke, editors (2002): *Automata, Logics, and Infinite Games, LNCS* 2500. Springer.

[10] A. Hordijk & A.A. Yushkevich (2002): *Blackwell Optimality*. In: E.A. Feinberg & A. Schwartz, editors: *Handbook of Markov Decision Processes,* chapter 8, Kluwer.

[11] D.A. Martin (1998): *The determinacy of Blackwell games. Journal of Symbolic Logic* 63(4), pp. 1565–1581.

[12] J.F. Mertens & A. Neyman (1981): *Stochastic Games*. *International Journal of Game Theory* 10, pp. 53–56.

[13] L. S. Shapley (1953): *Stochastic games. Proceedings Nat. Acad. of Science USA* 39, pp. 1095–1100.

[14] A.N. Shiryayev (1984): *Probability*. Springer.

[15] Daniel W. Stroock (2005): *An Introduction to Markov Processes*. Springer.